Molecular Biology of the Cell
Vol. 15, 1031–1043, March 2004

Expression Profiling of Mammalian Male Meiosis and Gametogenesis Identifies Novel Candidate Genes for Roles in the Regulation of Fertility[D]

Ulrich Schlecht,*† Philippe Demougin,*† Reinhold Koch,*† Leandro Hermida,† Christa Wiederkehr,† Patrick Descombes,‡ Charles Pineau,§ Bernard Jégou,§ and Michael Primig†‖

†Biozentrum and Swiss Institute of Bioinformatics, 4056 Basel; Switzerland; ‡Genomics Platform, National Center of Competence in Research Frontiers in Genetics, Université de Genève/CMU, 1211 Geneva 4, Switzerland; and §Group d'Etude de la Reproduction chez le Male-Institut National de la Santé et de la Recherche Médicale U. 435, Université de Rennes I, Campus de Beaulieu, 35042 Rennes cedex, Bretagne, France

Submitted October 24, 2003; Revised December 1, 2003; Accepted December 1, 2003
Monitoring Editor: Keith Yamamoto

We report a comprehensive large-scale expression profiling analysis of mammalian male germ cells undergoing mitotic growth, meiosis, and gametogenesis by using high-density oligonucleotide microarrays and highly enriched cell populations. Among 11,955 rat loci investigated, 1268 were identified as differentially transcribed in germ cells at subsequent developmental stages compared with total testis, somatic Sertoli cells as well as brain and skeletal muscle controls. The loci were organized into four expression clusters that correspond to somatic, mitotic, meiotic, and postmeiotic cell types. This work provides information about expression patterns of ~200 genes known to be important during male germ cell development. Approximately 40 of those are included in a group of 121 transcripts for which we report germ cell expression and lack of transcription in three somatic control cell types. Moreover, we demonstrate the testicular expression and transcriptional induction in mitotic, meiotic, and/or postmeiotic germ cells of 293 as yet uncharacterized transcripts, some of which are likely to encode factors involved in spermatogenesis and fertility. This group also contains potential germ cell-specific targets for innovative contraceptives. A graphical display of the data is conveniently accessible through the GermOnline database at http://www.germonline.org.

INTRODUCTION

During mammalian spermatogenesis, primordial germ cells develop into spermatogonia, giving rise to spermatocytes that undergo two meiotic divisions to become round spermatids. These cells differentiate into spermatozoa during spermiogenesis (Pineau *et al*., 1993a; Griswold, 1998). The development of male germ cells is governed by testicular Sertoli cells through secretion of, among others, signaling molecules and factors involved in cell adhesion (Griswold, 1998; Jégou *et al*., 1999). Genes important for male gametogenesis and fertility in the mouse include loci involved in highly conserved landmark events such as meiotic recombination (Keeney *et al*., 1997; Edelman *et al*., 1999; Luo *et al*., 1999; Kneitz *et al*., 2000; Romanienko and Camerini-Otero, 2000; Lipkin *et al*., 2002), formation of the synaptonemal complex (that holds the homologous chromosomes together), sister chromatid cohesion, and checkpoints and factors required for the meiotic cell cycle (Liu *et al*., 1998; Zickler and Kleckner, 1998; Wolgemuth *et al*., 2002; Petronczki *et al*., 2003). Other factors essential for postmeiotic stages of spermatogenesis include protein kinases (Miki *et al*., 2002), proteases (Adham *et al*., 1997; Shamsadin *et al*., 1999), chromatin condensation factors (Yu *et al*., 2000; Adham *et al*., 2001; Cho *et al*., 2001), and proteins involved in cell-cell adhesion, notably of Sertoli cells and spermatids (Mannan *et al*., 2003).

Genes required for spermatogenesis and fertility often encode transcripts present in meiotic or postmeiotic germ cells but not in somatic tissue (Eddy, 2002; McLean *et al*., 2002). The notion that developmental-stage or tissue-specific expression indicates an important function is supported by results from high-throughput gene inactivation studies (Rabitsch *et al*., 2001; Coaiacovo *et al*., 2002) based upon expression profiling data obtained with sporulating yeast cells (Chu *et al*., 1998; Primig *et al*., 2000) and germ cells in the worm (Reinke *et al*., 2000). Transcriptome studies have increased the pace at which genes important for sexual reproduction are identified in these important model systems (for review, see Schlecht and Primig, 2003). In this context it should be emphasized that budding yeast genes important for sporulation (a process analogous to spermatogenesis) were demonstrated to be transcriptionally induced during the process in a statistically significant manner (Deutschbauer *et al*., 2002). A number of very recent array profiling analyses have addressed the problem of gene expression in the mammalian male germ line. However, these

Article published online ahead of print. Mol. Biol. Cell 10.1091/mbc.E03–10–0762. Article and publication date are available at www.molbiolcell.org/cgi/doi/10.1091/mbc.E03–10–0762.

[D] Online version of this article contains supplementary material. Online version is available at www.molbiolcell.org.

* These authors contributed equally to this study.
‖ Corresponding author. E-mail address: michael.primig@unibas.ch.

experiments were either limited to postmeiotic germ cells (Ostermeier *et al*., 2002), solely based upon total testis samples (Sha *et al*., 2002; Sadate-Ngatchou *et al*., 2003; Schultz *et al*., 2003), or did not include somatic controls (Pang *et al*., 2003). Moreover, none of these studies provided a searchable database accessible via the internet.

We sought to explore gene expression in rat Sertoli and germ cells to provide clues to the mitotic, meiotic, and postmeiotic functions of many hundred uncharacterized transcripts. Among 11,955 loci analyzed using a statistical approach, we identified 1268 as being strongly differentially regulated in testicular somatic versus germ cells. Those loci were organized into four broad categories whose transcription levels peak in either somatic or germ cells at mitotic, meiotic, and postmeiotic stages of development. We included somatic testicular Sertoli cells as well as brain and skeletal muscle samples in our analysis as negative controls for transcriptional induction/expression in germ cells. The results clearly suggest that a substantial fraction of the known genes expressed in meiotic and postmeiotic germ cells but not in the examined somatic tissues are important for meiosis, spermatogenesis, and/or fertility. Finally, quantitative polymerase chain reaction (Q-PCR) assays were carried out to underscore the reliability and reproducibility of the microarray expression data. The results of this analysis constitute an extremely useful source of information about thousands of mammalian transcripts. Moreover, it specifically marks out >290 as yet unknown candidate loci as being potentially involved in spermatogenesis or fertility. Because many of these loci are preferentially or maybe even exclusively expressed in germ cells, they could encode targets for novel approaches to temporarily and reversibly inhibit male fertility.

The expression data as well as supplementary material, including analysis programs are accessible at http://www.bioz.unibas.ch/personal/primig/rat_spermatogenesis. The data can be searched at reXbase integrated into the *Rattus norvegicus* section of GermOnline (http://www.germonline.org), a novel cross-species community annotation knowledgebase (Primig *et al*., 2003; Wiederkehr *et al*., 2004a,b) that provides a graphical display of microarray expression signals relevant for germ cell differentiation.

MATERIALS AND METHODS

Cell Purification Protocols

Sertoli cells were isolated from eight rats at 20 d postpartum (dpp) as described previously (Toebosch *et al*., 1989). To isolate spermatogonia, testes of 25 Sprague-Dawley rats at 9 dpp were excised and decapsulated. Seminiferous epithelial cells were dispersed by collagenase treatment and separated at a purity of >90% as described previously (Guillaume *et al*., 2001). Pachytene spermatocytes and early spermatids were prepared by centrifugal elutriation to a purity >90% from eight rats at 90 dpp as described previously, except that cells were mechanically dispersed (Pineau *et al*., 1993b). Total testicular samples were produced by excising and snap freezing testes from three Sprague-Dawley rats at 90 dpp in liquid nitrogen. The outermost connective tissue capsule was surgically removed on the frozen organs before they were manually ground up using mortar and pestle. Total RNA was purified using the RNeasy kit (QIAGEN, Valenica, CA) following the instructions of the manufacturer. Tissue samples from brain (Lewis, 60 dpp) and skeletal muscle (Wistar, 70 dpp) for microarray analysis were isolated from adult rats according to standard procedures as recommended by the manufacturer (Affymetrix, Santa Clara, CA).

Q-PCR Analysis of mRNA Concentration

cDNA was synthesized using random hexamers and Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA) following the instructions of the manufacturer. The amplicons were designed with the program Primer Express 2.0 (Applied Biosystems, Foster City, CA) by using default parameters such that they spanned exon boundaries. The specificity of each primer pair as well as the efficiency of the amplification step were tested by assaying serial dilutions of cDNA oligonucleotides specific for target and normalization control loci and expressed sequence tag (EST) (see Web site). PCR reactions were carried out in triplicate by using an SDS 7900 HT instrument (Applied Biosystems). Raw Ct values obtained with SDS 2.0 (Applied Biosystems) were imported into Excel (Microsoft, Redmond, WA) to calculate the normalization factor and the fold changes with the geNorm script as published (Vandesompele *et al*., 2002).

cRNA Target Synthesis and GeneChip Hybridization

Fifty micrograms of total RNA was purified on average from the samples by using RNeasy Mini-Spin columns (QIAGEN) with standard protocols provided by the manufacturer. The cell pellets were resuspended in the proprietary buffer RLT that contains guanidine isothiocyanate and lysed by shearing in a 2-ml syringe. Then 600 μl of the supernatant was mixed with 600 μl of 70% ethanol, loaded onto an RNeasy column, and washed and eluted in 100 μl of distilled water. Total RNA quality was monitored by loading ~200 ng onto an RNA Nano 6000 Chip that was processed with the 2100 Bioanalyzer (Agilent, Palo Alto, CA). Biotin labeling of RNA was performed as described in the Expression Analysis Technical Manual (Affymetrix) with minor modifications as indicated below. To synthesize single-stranded cDNA, 13 μg of total RNA was mixed with oligo-dT and incubated with SuperScript II reverse transcriptase (Invitrogen) at 42°C for 1 h. After synthesis of the second cDNA strand by using the SuperScript double-stranded cDNA synthesis kit (Invitrogen), the material was extracted with phenol-chloroform-isoamyl alcohol, and precipitated with 0.5 volumes of 7.5 M ammonium acetate and 2.5 volumes of ethanol. Fifty percent of the cDNA was used for an in vitro transcription reaction by using the BioArray high yield RNA transcript labeling kit T7 (Enzo Diagnostics, New York, NY) to synthesize cRNA in the presence of biotin-conjugated UTP and CTP analogs. Approximately 50–100 μg of labeled cRNA from each reaction was purified by using RNeasy mini-spin columns, and roughly 300 ng was analyzed on RNA Nano 6000 Chips. The cRNA targets were incubated at 94°C for 35 min and the resulting fragments of 50–150 nucleotides were again monitored using the Bioanalyzer. All synthesis reactions were carried out using a PCR machine (T1 thermocycler; Biometra, Göttingen, Germany) to ensure the highest possible degree of temperature control.

The hybridization cocktail (220 μl) containing fragmented biotin-labeled target cRNA at a final concentration of 0.05 μg/μl was transferred into Rat U34A and U34B GeneChips (Affymetrix) and incubated at 45°C on a rotator in a hybridization oven 640 (Affymetrix) for 16 h at 60 rpm. The arrays were washed and stained by using a streptavidine-phycoerythrin (SAPE) conjugate (Molecular Probes, Eugene, OR). To increase the signal strength the antibody amplification protocol was used (EukGE-WS2v4; Affymetrix Expression Analysis Manual). The GeneChips were processed with a GeneArray Scanner (Agilent) by using the current default settings. DAT image files of the microarrays were generated using Microarray Analysis Suite 5.0 (MAS; Affymetrix).

Data and Cluster Analysis

CEL data files were computed using the statistical algorithm implemented in MAS 5.0 (http://www.affymetrix.com/support/technical/technotes/statistical_algorithms_technote.pdf). The data were further analyzed using programs developed in R, a programming language and developer environment for statistical computing and graphics (http://www.r-project.org). The Robust Multi-array Analysis algorithm as implemented in the BioConductor package was employed for data normalization using the quantile method as provided in the package (http://bioconductor.org). This method aims at making the distribution of probe intensities the same across all arrays in a given set (Bolstad *et al*., 2003; Irizarry *et al*., 2003b). Gene expression signal calculation was based upon the Perfect Match values from each probe set as published previously (Irizarry *et al*., 2003a). The program written for data preparation and analysis in this study is available at our web site (follow the link "download software" in the navigation bar).

To identify differentially expressed genes, the F statistic was used to compute the variations between replicates as compared with those observed between two different samples. Based upon this method, a complete permutation test was carried out such that the expression changes observed when true replicates were analyzed, were compared with the changes that occur when any two samples were considered to be replicates. This approach helps distinguish true signal changes from those that occur by chance. Only the genes for which a p value <0.05 and a SD of >1 was calculated were included in the list of differentially expressed genes for further analysis by clustering.

Dendrograms of the expression patterns and samples were computed by hierarchical clustering using Ward's Minimum Variance and Complete Linkage Methods, respectively, as implemented in R in the hclust function from the *mva* package. In both cases a Euclidian distance measure was used. Dendrograms were combined with a heat map displaying color-coded expression signal intensities for each gene.

To subgroup the genes according to their overall transcription patterns but not signal intensities the values for each gene were centered on zero and scaled to a SD of 1. The transformed values were then analyzed using partitioning around medoids (PAM) (Kaufman and Rousseeuw, 1990) and k-means clustering as implemented in R and GeneSpring 5.1 (Silicon Genetics,

Table 1. Cell purification efficiency and sample quality

Testicular cell type	Isolation procedure	Purity (%)	Contaminating material	Reference(s)
Sertoli cell	Sequential enzymatic digestion	96	Peritubular cells (<2%), Mitotic and meiotic germ cells (2%)	Pineau *et al.* (1993b)
Spermatogonium	Sequential enzymatic digestion + sedimentation at unit gravity	85	Fragments of Sertoli cell cytoplasm, peritubular cells, Leydig cells	Guillaume *et al.* (2001)
Pachytene spermatocyte	Mechanical dispersion + centrifugal elutriation	90	Early spermatids	Pineau *et al.* (1993b)
Early spermatid	Mechanical dispersion + centrifugal elutriation	90	Early spermatids, heads from elongated spermatids, spermatid residual bodies, spermatocytes	Pineau *et al.* (1993b)

Redwood City, CA), respectively. The significance of PAM results after several rounds of analysis using increasing numbers of clusters was verified by using silhouette plots. The silhouette values are a measure of the degree of similarity of expression patterns in a given cluster compared with all other patterns in the other clusters. Both PAM and silhouette are part of the cluster package in R.

Similarity Search

The UniGene Rn.data.gz file was parsed and read into a MySQL relational database. The fields from the Rn.data file used were as follows: ID (UniGene cluster ID), TITLE (title for the UniGene cluster), GENE (gene symbol), PROTSIM (protein similarity/homology data of rat UniGene clusters to loci in rat and other model organisms), and SEQUENCE (sequence information of all sequences that belong to each UniGene cluster).

We generated the following probe set annotation groups: 1) probe sets not in UniGene 118, 2) probe sets that represent annotated genes, 3) probe sets that represent genes not yet annotated with at least one homolog, and 4) probe sets which represent genes not yet annotated with no homologues and a significant protein-NR BLAST alignment. To determine whether a probe set was in UniGene 118 and which UniGene cluster it was in, the source accessions for each probe set were mapped to UniGene 118 clusters. UniGene gene symbol and BioConductor PubMed ID information were used to determine whether a probe set represented an annotated gene. UniGene PROTSIM information was used to determine whether the gene represented by a probe set had a homolog. Probe sets corresponding to genes not yet annotated with no homologues were aligned to National Center for Biotechnology Information's protein-NR database by using the BLASTx program with a minimum e-value of 1e-5.

Compliance

Both CEL and CHP data files corresponding to the Sertoli (SE1 and 2), germ cell (SG, SC, and ST1 and 2, respectively), and total testis samples (TT1 and 2) were uploaded to the European Bioinformatics Institute's ArrayExpress public data repository at http://www.ebi.ac.uk/arrayexpress/(Accession number E-MEXP-31). CEL feature level data files of all samples (including brain [BR] and skeletal muscle [SM]), and Q-PCR raw data are available at our Web site.

RESULTS

Experimental Design

The aim of this large-scale microarray study was to identify genes that are expressed during the mitotic, meiotic, and postmeiotic developmental stages of mammalian germ cells and to further characterize their transcription patterns in total testis samples and somatic control cells or tissues. To this end we have used high density oligonucleotide microarrays (U34 A and B GeneChips) from Affymetrix. The transcript levels of 11,955 rat loci and ESTs were simultaneously measured in highly enriched testicular germ and Sertoli cell populations (Table 1) as well as in brain and skeletal muscle tissue. Because RNA degradation causes data to be less reproducible and makes it more difficult to detect low-abundance transcripts, we included a sample prepared from total testis by using a rapid extraction protocol as a control for total RNA quality. All total RNA (Figure 1, a and c) and cRNA target molecule preparations (b and d) made from purified cells and total testis tissue are of very high quality.

Identification of Differentially Expressed Transcripts in Somatic versus Germ Cells

The terms *locus* and *gene* are used as synonyms throughout the text, whereas *probe set* refers to a row of oligonucleotides present on the microarray that are complementary to one transcript or one EST. Note that the arrays can detect more than one transcript/EST for one gene; that is to say a given number of probe sets detects an equal number of transcripts/ESTs that ultimately correspond to a smaller number of actual genes. Gene filtration was carried out as follows: the PerfectMatch values were corrected for background noise, normalized, and summarized to one value for each probe set by using methods implemented in the Robust Multi-Array Analysis package (see MATERIAL AND METHODS for more details). This innovative approach does not take the values obtained with the Mismatch Oligonucleotide into account. This increases the sensitivity of the measurements because the Mismatch Oligonucleotide signal is thought not just to reflect nonspecific hybridization events but also to contain true expression information that is lost when the default statistical algorithm for data analysis provided by the manufacturer is applied (Bolstad *et al*., 2003; Irizarry *et al*., 2003a). Permutation tests based upon the F statistic that do not depend upon normal distribution of the data were used to identify 9864 probe sets as being differentially expressed within the cell populations analyzed (Clarke and Cooke, 1998). In these cases, the null hypotheses (all experimental conditions yield the same mean expression value for a given gene) were rejected at the 5% significance level. Using an approach by Storey, 2002, it was estimated that among the 9864 probe sets should be 75 false positives. Finally, a subset of 1508 probe sets for which a p value of <0.05 was computed and whose expression signals displayed a SD >1 across the 10 data sets (hybridization experiments) was selected for further analysis. It should be emphasized that this approach yields a conservative estimate of the number of differentially expressed genes. The signals obtained in duplicate were found to be highly reproducible for both U34 A and B GeneChips as summarized in the box plots shown in Figure 2a. The scatter plot matrix of 1508 probe sets also indicates that expression levels are highly similar within replicates, whereas they split up into four broad clusters of differentially expressed loci (shown in color) when different testicular germ and Sertoli cell popu-

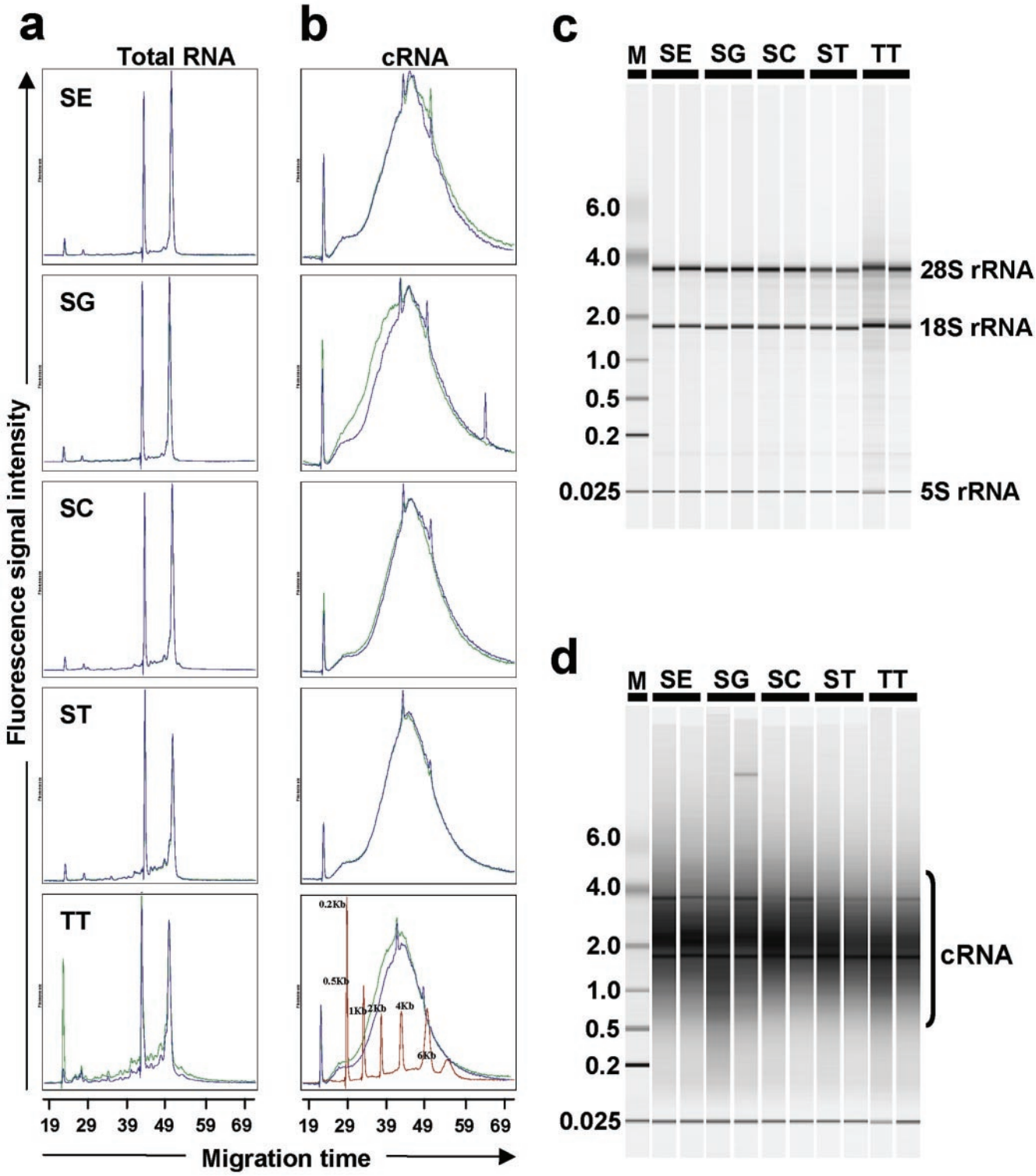


Figure 1. Total RNA and cRNA target quality assessment. Purified total RNA from Sertoli cells (SE), spermatogonia (SG), spermatocytes (SC), spermatids (ST), and total testis (TT) samples were analyzed for their concentration and overall length by using the Agilent Bioanalyzer RNA Chip. Fluorescence intensities measured with total RNA and cRNA target molecules are plotted against migration time in seconds as indicated in a for total RNA and b for cRNA. Signals obtained with identical replicates are shown in green and blue graphs, respectively. The peaks corresponding to the molecular weight RNA ladder are displayed in the total testicle cRNA plot. Virtual gels of the total RNA and cRNA samples as indicated are shown in c and d, respectively. M, molecular weight markers. rRNAs and the cRNA ladder are indicated.

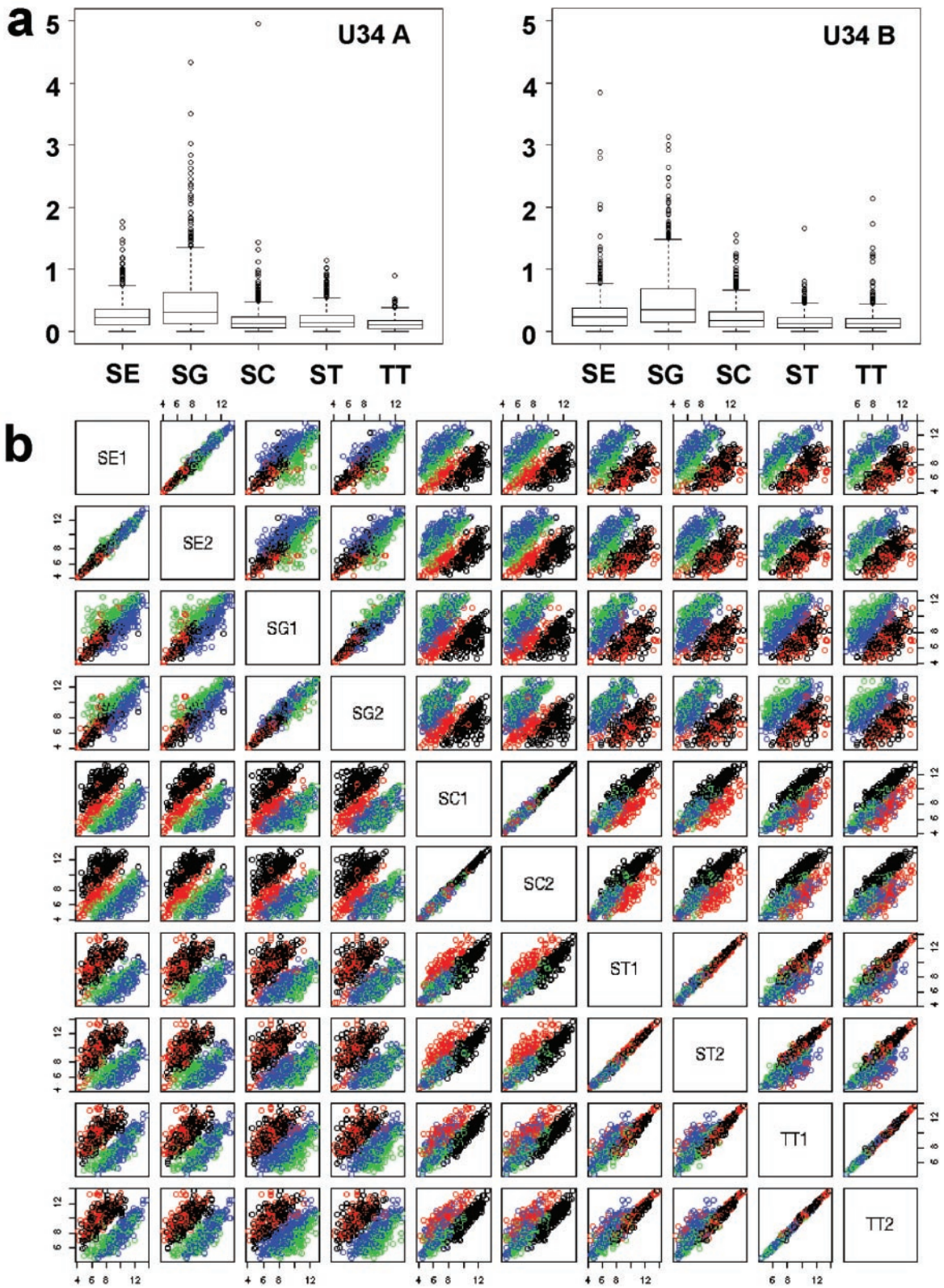


Figure 2. Comparing gene expression levels of 1508 probe sets within replicates and different cell populations. Box plots in a display the overall distribution of observed changes in signal intensities obtained with two replicates from five samples hybridized to the U34 A and B GeneChips as indicated. Shown are the absolute log2 fold change values. The features of clear outliers were manually inspected on the corresponding DAT image file to exclude artifacts. The scatter plot matrix in b summarizes a systematic comparison of the log2-scaled expression signals from 1508 transcripts among all samples as indicated. Genes that fall into the somatic, mitotic, meiotic, and postmeiotic clusters are symbolized by circles colored in blue, green, black, and red, respectively.

lations are compared (Figure 2b; brain and muscle data are shown in Web Figure 1).

Identification of Transcripts Represented by Probe Sets through Similarity Searches

The RG-U34 A and B GeneChips contain a total of 17,590 probe sets that correspond to 11,955 UniGene clusters, each of which is a group of sequences that belong to one unique gene; 1508 probe sets where found to display strongly changing expression signals during rat spermatogenesis (including 12 controls for rat GAPDH and Hexokinase that were not further analyzed). Because RG-U34 is based on Rat UniGene Build 34, we sought to identify the number of currently annotated loci represented by the 1496 probe sets, by using Build 118, which contains 53,186 distinct clusters.

The search revealed that the 1496 probe sets correspond to 1268 UniGene clusters of which 592 were annotated (Figure 3). This group comprises 487 loci that were allocated a gene symbol in the UniGene annotation and 105 that were manually processed by us. The group of annotated loci includes some that are similar to yeast genes important for the mitotic and meiotic cell cycle (Cdc42/*CDC42*, Ccnb1/*CLB4*, Ccnd2/*CLB6*, Stk6/*IPL1*, and Tesk2/*TPK3*) and the regulation of meiosis (Mak/*IME2* and Gsk3β/*RIM11*) (Sicinski *et al*., 1996; Gromoll *et al*., 1997; Rosok *et al*., 1999; Goepfert *et al*., 2002; Shinkai *et al*., 2002; Guo *et al*., 2003; Lui *et al*., 2003). See SGD (http://yeastgenome.org) for references of the yeast genes. Among the 676 loci not yet annotated are 383 genes that have potential homologues in other species, notably *Saccharomyces cerevisiae* (provided by HomoloGene). Examples of potentially important uncharacterized rat transcripts (ESTs) include AA874887 and AI012651 that are similar to yeast genes involved in sister chromatid cohesion (*SMC4*) (Freeman *et al*., 2000) and mitotic chromosome condensation (*BRN1*) (Ouspenski *et al*., 2000), respectively (see reXbase for more details). Finally, we identified 62 transcripts that may have homologues not previously identified by HomoloGene and 231 transcripts expressed in germ cells that did not match any other currently known sequences in the comprehensive National Center for Biotechnology Information Protein NR database (Figure 3).

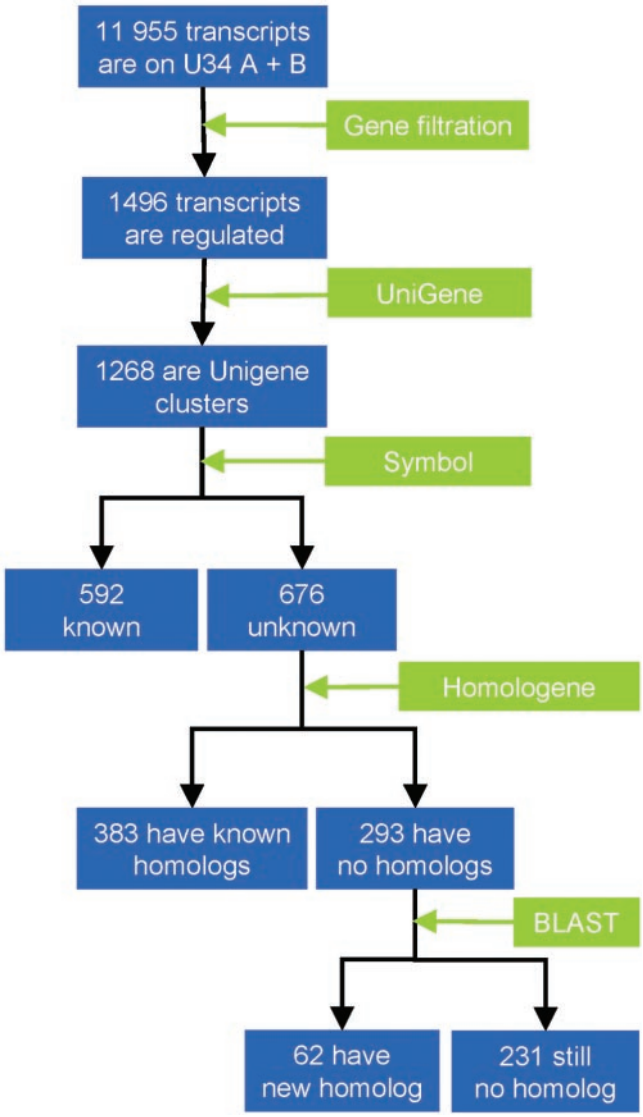


Figure 3. Mining U34 A and B GeneChips for annotated and novel transcripts. This figure summarizes the steps taken to identify among all transcripts represented on the GeneChips those that are differentially expressed (Gene filtration; among the 1508 initial transcripts 12 were removed because they were hybridization controls), those that were in UniGene Build 118 (UniGene), those that are encoded by known genes (defined as bearing a genetic name; Symbol), those that had known homologues (Homologene), and finally those that were not known to have a homologue or those that do not seem to be similar to any currently annotated protein (BLAST).

Clustering Gene Expression Profiles in Somatic and Germ Cells

To verify the clustering behavior of replicated and distinct samples and to get a broad overview of the timing of gene induction, we first explored the data set by hierarchical clustering of the 10 samples and 1508 expression patterns. The results are displayed by dendrograms (trees) and a color-coded heat map. As expected, replicates clustered together in all cases (top dendrogram in Figure 4). Furthermore, we observed clustering of samples from Sertoli cells and spermatogonia, indicating that differentiated somatic and mitotically growing spermatogonia share a substantial set of similarly expressed/induced genes. In contrast to that, meiotic and postmeiotic spermatocytes and spermatids are grouped together with adult total testis samples (top dendrogram). The data indicate that four broad patterns exist in purified Sertoli and germ cells that can be summarized on the basis of differential expression in somatic/mitotic versus meiotic/postmeiotic samples, respectively (left dendrogram and indications in Figure 4).

To try and identify the minimal set of clusters that represent the observed expression signals and to assign biological functions to groups of genes whose expression correlates with cell type (Sertoli cells), mitotic growth (spermatogonia), or meiotic developmental stage (spermatocytes and spermatids), the PAM algorithm was used. Using centered and scaled data to minimize the impact of expression levels, we defined four typical transcriptional patterns (the medoids) within the group of 1508 transcripts (Figure 5a) and then grouped all patterns around them. A graphical display of the expression patterns was produced using GeneSpring 5.1 (Figure 5b). These classes were verified by comparing the degree of similarity of a given pattern to those within its own cluster and to those in all other clusters by using silhouette plots (Figure 5c). Note that the somatic and mitotic expression clusters do not necessarily indicate specific expression but group genes together that display peak transcription in either of the cell types. The results were verified

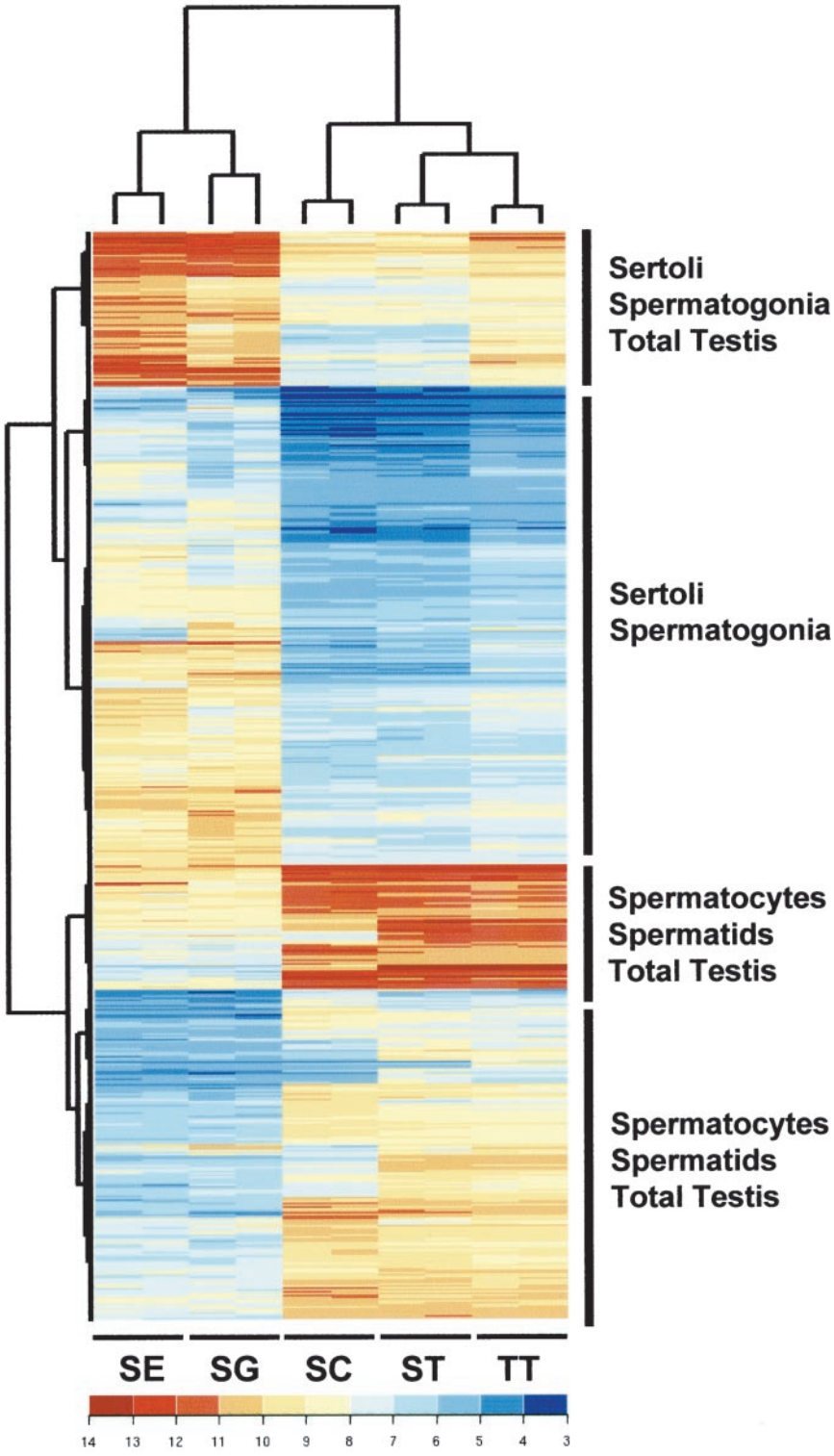


Figure 4. Hierarchical clustering of 10 samples and 1508 probe sets. A heatmap of 1508 transcripts and two dendrograms that group genes (left) and samples (top) together are shown. The samples are arranged such that somatic control Sertoli cells (SE) are followed by mitotic spermatogonia (SG), meiotic (pachytene) spermatocytes (SC), postmeiotic early spermatids (ST), and total testis (TT) tissue. Each line is a gene and each column is a sample. Clustered genes and samples as well as the color code for expression levels are given. Expression signal intensities are shown in red and blue indicating high and low expression, respectively. The cells displaying peak expression levels for a given group marked by a black bar are indicated.

using the k-means clustering algorithm implemented in GeneSpring 5.1 (our unpublished data).

Annotation data and references for all genes described below are summarized in Table 2 (see Table 2 in the supplementary material and on the web site). In addition to that, detailed curated descriptions of meiotically induced/expressed genes identified in this study will be available through the locus report pages of GermOnline (http://www.germonline.org).

The somatic expression class contains 474 transcripts that are predominantly (but often not exclusively) expressed in Sertoli cells from adult rats. This group also contains ~220 unknown transcripts. Among the annotated loci, ~100 genes have previously been demonstrated to be expressed in testes

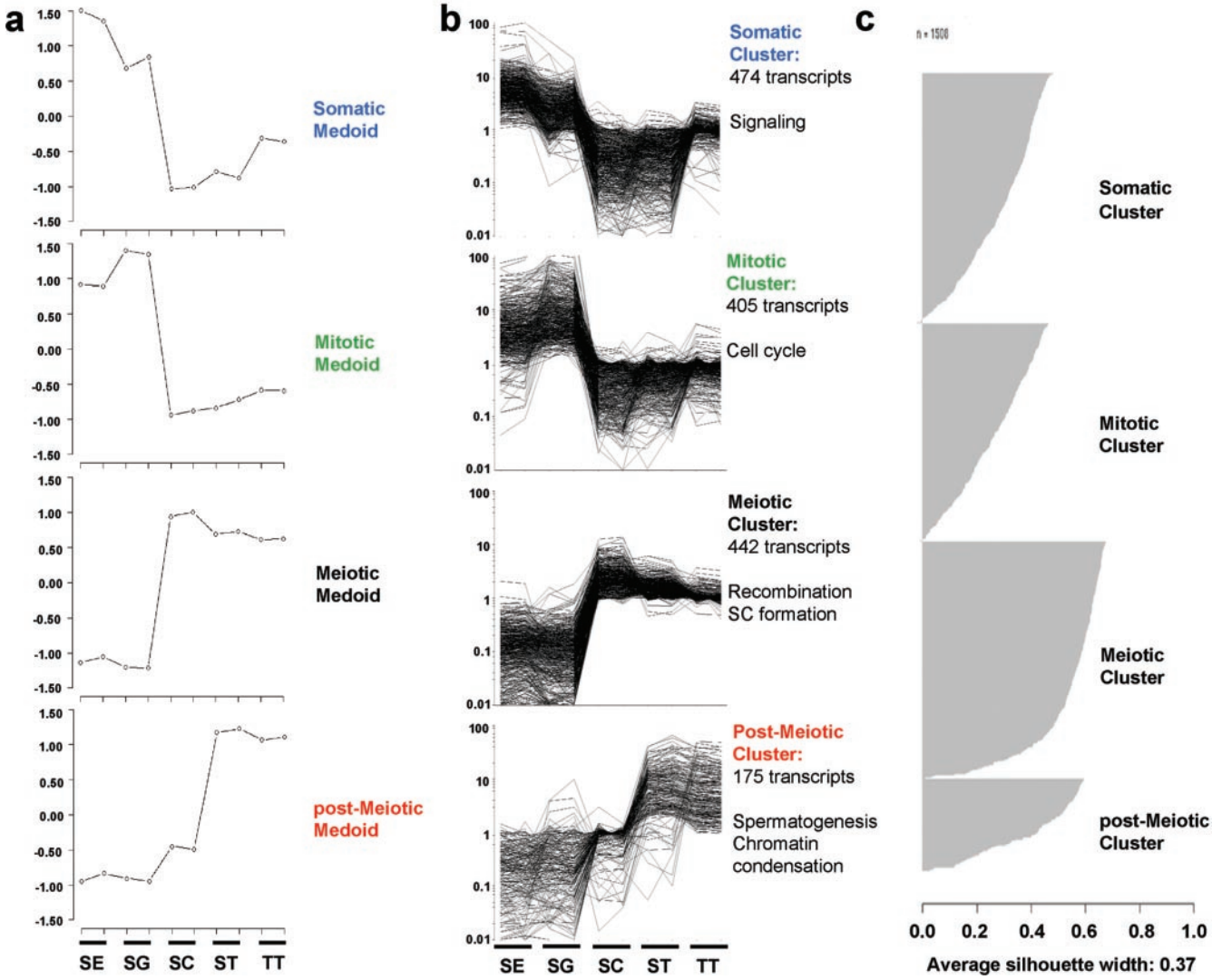
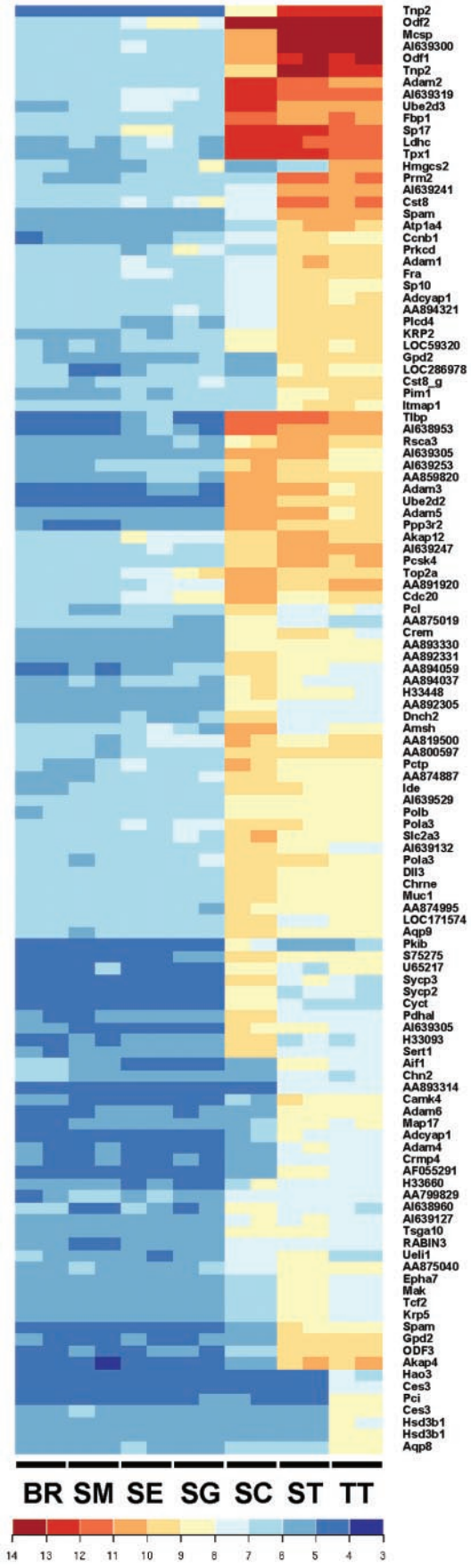


Figure 5. PAM clustering of 1508 probe sets. (a) Medoids representing the four clusters covering the samples as indicated. (b) Log-scale graphical display of the expression patterns within the somatic, mitotic, meiotic, and the post-meiotic clusters as determined by PAM. For visualization, the data computed with MAS 5.0 were scaled and normalized using the default settings of GeneSpring 5.1. Silhouette plots of the clusters are displayed in c. The scale of the silhouette is indicated and the average width is given at the bottom of the panel.

and/or to play roles in testicular and Sertoli cell functions. They are involved for example in energy and fatty acid metabolism (Cycs, Ldhb, Scd2, and Scp2), ubiquitin-mediated proteolysis (Ube2d3 and Uchl1), oxidative stress and serum response (c-Fos, Prdx1, Prdx2, Sod3, Gsta1, Gstm2, and Gstp2), and cell-cell signaling/interaction (Ctsl, Il6st, Il6r, Nr0b1, Psap, Shbg, and LOC286916 [Testin]). We also found cell adhesion factors (Cdh2 and Ctnnb1) and confirmed expression in germ cells of genes involved in tumor formation/gonadogenesis (Wt1) and X-Chromosome inactivation (Tsx). The mouse homologue of Tsx was demonstrated to be transcribed in mitotic germ cells during puberty as well as in adult Sertoli cells; this pattern was correctly detected by microarray analysis (Cunningham *et al*., 1998). Finally, we find expression in Sertoli cells and spermatogonia of a cell surface glycoprotein (Cd9) previously demonstrated to be expressed in spermatogonial stem cells (Kanatsu-Shinohara *et al*., 2003). Cd9 is required for sperm–egg interaction and female fertility (Le Naour *et al*., 2000).

The somatic class also contains >150 transcripts for known genes that were characterized in other tissues. Notably, we detect factors involved in cell cycle and growth control (Cdc42, Gas7, Rbbp7, and Ywhaβ), transcriptional regulation (Ssdp and Ureb1), signal transduction (Grb14, Jak1, Ptgfr, and Rap1b), and protein degradation (Adam10, Nedd4a, Prss11, Psma2, Psma3, Psme1, and Psme2). Another large group consists of factors involved in energy metabolism (Aldoa, Aldr1, Atp1a1, Atp2c1, Atp5g2, Mor1, Pgam1, and Pkm2). A gene encoding a factor involved in vitamin A transport (Rbp1) was found to be highly expressed both in Sertoli cells and spermatogonia. This confirms previous reports showing that vitamin A (retinol) regulates a variety of testicular functions in rodents (Livera *et al*., 2002). Interestingly, a factor involved in neurodegenerative conditions (Prnp, encoding a major prion protein precursor) was shown to be expressed in Sertoli cells and spermatogonia. This is in keeping with the fact that the Prnp homologue Dpl regulates male mouse fertility (Behrens *et al*., 2002).



The mitotic expression class consists of 405 transcripts, including >200 uncharacterized mRNAs, which are up-regulated in mitotically growing spermatogonia. Among ~60 loci that were associated with testicular expression and/or spermatogonia in the literature, we detect the messages of genes required for cell cycle regulation (Ccnd2), components of the extracellular matrix (Col3a1 and Mgp), hormone signal transduction (Ctgf, Egr1, Fgfr1, Igfbp2, Igfbp3, Pdgfa, and Vegf), and serum response and transcriptional regulation (Jun, JunB, Id2, Klf9, Stat3, and Zfp36).

Approximately 140 genes that were not known to be expressed in germ cells include transcription factors (Fosl1, Copeb, and Gata4), a protein phosphatase 2a catalytic subunit (Ppp2ca), and an AP-1–regulated serum-inducible protein that is involved in inflammatory processes in the brain and pancreas (Scya2). The latter may play a role in the testicular immune response to viral infections (Le Goffic *et al*., 2002). The expression level of about one-half of the genes in the mitotic cluster specifically peaks in mitotically growing spermatogonia and then substantially decreases at later meiotic stages. This is the case with, e.g., histones (H1f0 and H3f3b), ribosomal proteins (Rpl35, Rps3, and Rps4), and motor proteins (Mrlcb, Myh9, Mylc2a, Tpm1, and Tpm4).

The meiotic expression class contains 442 transcripts highly induced in spermatocytes. Transcripts that fall into this class, which includes 280 ESTs, are either only transiently up-regulated during meiosis or they continue to be expressed postmeiotically after meiotic induction. As expected, we found genes in this group that are involved in SC formation (Sycp2 and Sycp3), DNA repair (Polβ), and chromatin condensation (Top2a). Moreover, we find metalloproteases (Adam2, Adam3, and Adam5), factors necessary for ubiquitin-mediated protein degradation (Ube2d2 and Psmc3), transcriptional regulators (Crem and Miz1), enzymes involved in energy metabolism whose somatic isoforms are expressed in Sertoli cells (Cyct and Ldhc; see also somatic cluster), and a component of the flagellum ultimately required during spermiogenesis (Odf2). Previously undetected expression in germ cells was observed for a factor originally identified in Sertoli cells (Sert1), a gene involved in cell cycle regulation (Cdc20) and transcription factors involved in organogenesis of the heart, gut, and lung, for example (Pitx2 and Hipk3).

The postmeiotic expression class contains 175 transcripts, including 70 that encode uncharacterized mRNAs. The transcripts in this class are detected predominantly in spermatids and, in almost all cases, in the total testis sample. The latter provides not only a control for mRNA integrity but also a source for transcripts present in differentiated sperm that constitute approximately two-thirds of the testicular cell mass. This class contains almost 60 known genes that encode proteins implicated in regulation of meiosis (Gsk3β), cell cycle control (Ccnb1), protein phosphorylation (Clk3 and Mak), and degradation (Adam1, Adam4, Adam6, and Cst8), motor proteins (Krp2 and Krp5), water permeability (Aqp7 and Aqp8), and cAMP signaling (Akap1). Other examples include factors involved in DNA condensation during sperm differentiation (Prm2, Prm3, Tnp1, and Tnp2), sperm maturation (Pp2a, Odf1, and Odf3), adhesion (Spam and Sp10), and motility (Akap4).

Figure 6. Clustering of somatic control cells versus germ cells. A heatmap of the expression data obtained with somatic controls from brain (BR) and skeletal muscle (SM) are compared with Sertoli and germ cells as indicated. The names of the corresponding loci and ESTs are given. Expression signal intensities are color coded as indicated.

Among known genes from somatic or ovarian tissues loci that play roles in inflammation (Aif1), signaling (Epha7), and an estrogen-regulated factor implicated in female reproduction (Itmap1) were identified. Our results indicate that these genes may play a broader role in fertility than previously thought.

Several loci display significant expression signals only in the total testis extract but not in any of the purified cell populations. The corresponding mRNAs could be specific for testicular cell types that were not analyzed in this study. This includes the steroidogenic Leydig cells that express, e.g., Cyp17, involved in the production of corticosteroids and androgens (for review, see Payne, 1990) or Hsd3b1 required for steroid synthesis (O'Shaughnessy *et al*., 2002).

Somatic Tissue Profiling of Germ Cell Genes

To further investigate the transcriptional patterns of transcripts highly expressed in germ cells, we included brain and skeletal muscle samples in the analysis of a selection of 617 cases from the meiotic and postmeiotic clusters (Figure 5b). This approach marks out 121 transcripts as being expressed in meiotic germ cells but neither in somatic Sertoli, nor in brain, nor in skeletal muscle cells (Figure 6). The group comprises genes that are known to be involved in testicular transcription (Crem), cell cycle regulation (Cdc20 and Ccnb1), synaptonemal complex formation (Sycp2 and Sycp3), DNA repair (Polβ), chromatin condensation (Top2a, Tnp1, and Tnp2), cell-cell contact (Adam1, 2, 4, 5, and 6; Tpx1; and Spam), proteolysis (Pcsk4), and water transport (Aqp8 and Aqp9). Importantly, this approach identified 37 ESTs displaying expression signatures that are concordant with possible roles for their corresponding genes in spermatogenesis and fertility.

Array Data Validation by Q-PCR

To validate the microarray expression signals we carried out Q-PCR assays and observed a very good correlation between both techniques in cases of three known genes and five ESTs (Figure 7): Gsk3β, an orthologue of the yeast *RIM11* kinase required for meiosis and spore formation (Malathi *et al*., 1997), was very recently shown to be expressed in testes where it regulates mammalian meiosis (Guo *et al*., 2003). The locus encoding the rat TATA box binding protein (Tbp, represented by the EST AA891928 that maps to NW_043264) also displayed the expected highly reproducible pattern of transcriptional induction in germ cells as reported previously (Persengiev *et al*., 1996). Aqp8 was claimed to be expressed in germ cells at various stages of development (Ishibashi *et al*., 2002). However, we detected significant expression in any of the purified cell populations neither by microarrays nor by Q-PCR. As opposed to that, a clear signal was observed in the total testicular extract (Figure 7). The ESTs whose germ cell expression was validated are AI030059 (no known homologue), AI639326 (similar to the human protein pir:T17234), AI639231 (similar to the human protein pir:T46480), and AI639319 (similar to the mouse locus NP_080924.1). Furthermore, we verified AI013870, which maps to the rat locus AY241457.1 that encodes a conserved histone acetyl transferase (see DISCUSSION).

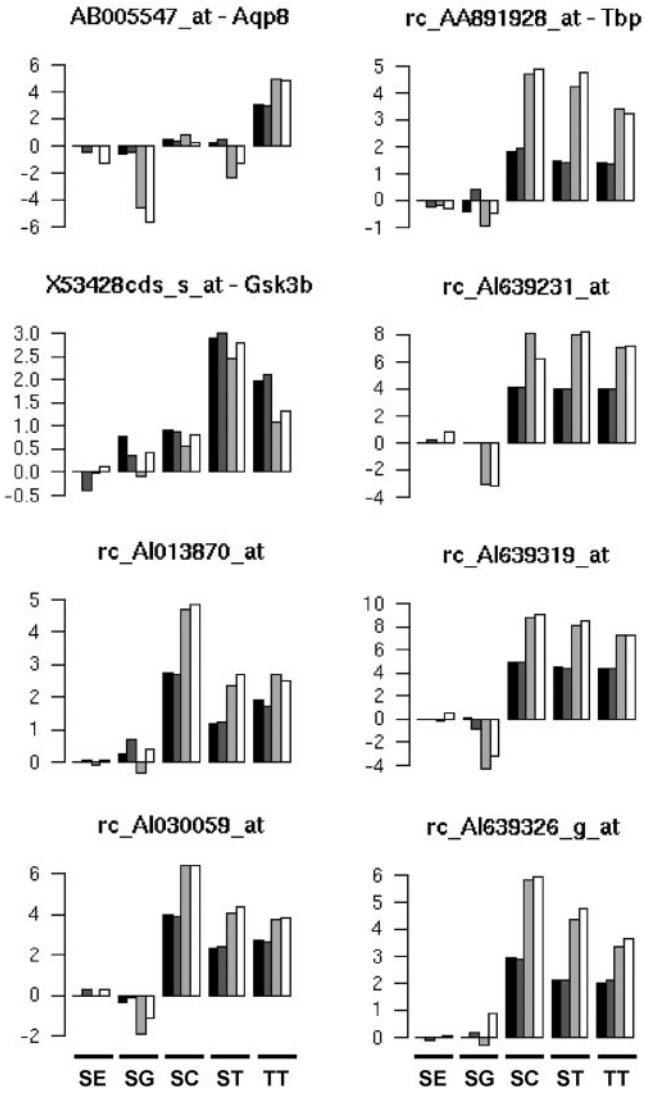


Figure 7. Q-PCR validation of selected transcripts. A graphical display of relative RNA concentrations as determined by quantitative PCR and microarray expression signals is shown. The samples are displayed in duplicate for array data (black and dark gray bars) and Q-PCR data (light gray and white bars; fold change relative to the SE sample normalized to 1) as indicated. Signals are given on a log2 scale.

DISCUSSION

Mining Mammalian Meiosis

This large-scale microarray experiment explores the patterns of transcriptional regulation underlying male meiosis and gametogenesis compared with testicular and nontesticular somatic control tissues in the rat, and it thereby confirms and substantially extends previous work in human and mouse (Ostermeier *et al*., 2002; Sha *et al*., 2002; Pang *et al*., 2003; Schultz *et al*., 2003). Results obtained in this study are based upon the analysis of ~35% of the genes expected to be present in a rodent genome (Waterston *et al*., 2002). It therefore provides a highly valuable data set readily available for the community via the GermOnline knowledgebase and constitutes a prototype approach for comprehensive transcriptomics in the field of eukaryotic sexual reproduction. It was our aim to identify loci that display strong expression throughout meiotic and postmeiotic stages of spermatogenesis because these genes are likely to play important roles in male gametogenesis and fertility. To this end, we have partially reconstituted the mitotic, meiotic, and the postmeiotic phases of rat sperm development by using highly enriched

spermatogonia, spermatocytes, and spermatids. We included purified somatic Sertoli cells, brain and skeletal muscle tissues, and total testis samples as controls for germ cell expression/induction and mRNA stability. Among 11,955 loci analyzed, 1268 showed statistically valid and strong differences in expression levels during the mitotic, meiotic, and postmeiotic phases of spermatogenesis as compared with Sertoli cells. Note that genes displaying more subtle changes of expression (that may still be biologically significant) were not necessarily included in the group of differentially expressed genes. This study confirms the transcriptional patterns of numerous previously detected loci and provides clues about the possible roles in male meiosis and gametogenesis for 293 as yet uncharacterized transcripts present in germ cells.

The 1508 probe sets (which correspond to 1268 loci) were explored and organized into four expression categories based upon the cell type and timing of induction during mitotic, meiotic, and postmeiotic stages of spermatogenesis by using hierarchical and PAM clustering algorithms. In the case of the latter, we focused on relative differential transcription patterns (by centering and scaling the data; see MATERIALS AND METHODS), rather than trying to categorize loci based upon expression levels that would have produced a larger number of clusters. This was due to the nature of our experimental approach because it is likely that mRNA concentrations change to a certain extent during germ cell purification compared with the in vivo level. The expression signals may therefore in some cases not accurately reflect the physiological promoter activity.

We demonstrate the reproducibility of array data by quantitative PCR, which consistently matches microarray hybridization signals. Correlation coefficients of the expression profiles determined for eight cases were found to vary between 0.83 and 0.99 (see correlation matrix in Web Figure 2). The profiling data reported here confirm Gsk3β expression in Sertoli and germ cells observed by RNA in situ hybridization. Inhibition of the Gsk3β kinase, a homologue of yeast Rim11 that phosphorylates Ume6 and Ime1 (Bowdish *et al*., 1994; Malathi *et al*., 1997), was shown to prevent premeiotic DNA replication in cultured rat germ cells (Guo *et al*., 2003). This gene is therefore a prototype case where increased mRNA concentration is concomitant to an elevated level of a protein that fulfills an important function in germ cells. The gene encoding Tpb (TATA binding protein) was demonstrated to be highly induced in meiotic and postmeiotic rat and mouse male germ cells compared with somatic tissue (Persengiev *et al*., 1996); both microarrays and Q-PCR analysis clearly reproduce this pattern, further underlining the reliability of the array data. Although we cannot confirm an earlier report claiming that Aqp8 is expressed in germ cells at various stages of development, our results are consistent with testicular expression and lack thereof in brain and skeletal muscle (Ishibashi *et al*., 1997). Moreover, although the in situ hybridization signal published by Ishibashi *et al.* (1997) is indeed localized within the seminiferous tubules, poor resolution of the image presented makes it difficult, if not impossible, to verify germ cell expression. An EST present on the U34 B array (AI013870) lead us to identify the rat homologue of a conserved putative histone deacetylase (HDAC; AY241457.1) as being strongly expressed in spermatocytes. The yeast homolog of the rat HDAC (*ESA1*) is required for mitotic growth (Smith *et al*., 1998) and Mof, its putative fly homolog, is involved in X-chromosome gene dosage compensation. This raises the intriguing possibility that the rat gene may play similar roles during spermatogenesis (Akhtar and Becker, 2000).

Experimental Design and Data Analysis

Is it possible to reconstitute a complex developmental pathway by using purified cell samples? Although it is likely that lengthy cell purification procedures affect transcript concentrations, there are several arguments in favor of profiling experiments based upon enriched germ cells. First, microarrays confirm numerous patterns from previously studied genes whose expression was analyzed by in situ hybridization in the natural cellular environment (e.g., Tsx and Gsk3β) (Cunningham *et al*., 1998; Guo *et al*., 2003). Second, the timing of induction of many meiotic and postmeiotic transcripts reported here correlates well with the time of function during spermatogenesis (e.g., Sycp2 and Sycp3, which encode SC components expressed in spermatocytes) (Offenberg *et al*., 1998; Schalk *et al*., 1998). Third, it is a feature of germ cells to contain mRNAs that are stored after transcription for translation at later stages; these mRNA species can therefore be detected by microarrays rather efficiently (for review, see Steger, 2001). In addition to that, signals from purified meiotic and postmeiotic germ cells were for many genes at least as strong (or even stronger) than those obtained with total testicle tissue (e.g., Ccnb1, Crem, Adam4, Adam6, and Akap4; see GermOnline). This indicates that mRNA degradation in purified cells is not a critical problem in most cases (Figure 6).

Our results are in keeping with previous findings that meiotic transcripts even from very early zygotic stages of embryological development are very stable and clearly detectable by microarrays in human sperm (Ostermeier *et al*., 2002). Moreover, similar experiments using mouse total testis as well as enriched cell samples also yielded abundant evidence that male germ cells are an excellent source for transcripts expressed/induced during spermatogenesis (Pang *et al*., 2003; Schultz *et al*., 2003). A direct comparison of our results with these studies is complicated by the fact that either different array technologies or utterly distinct analysis approaches were used. However, it is noteworthy that among 22 conserved genes important for spermatogenesis that are represented on mouse and rat GeneChips, 20 were identified by us and Schultz *et al.* (2003) as being up-regulated in male germ cells (Schultz *et al*., 2003). A better understanding of the conserved mammalian transcriptome of spermatogenesis will have to await a genome-wide comparative profiling study in mouse, rat and human; it should be mentioned that the array technology required for such an experiment is already available.

Cross-contamination of purified cell populations is likely to be irrelevant for most transcripts identified in this study because many reference genes known to be very strongly expressed in Sertoli cells, spermatocytes, or spermatids (e.g., Tsx, Sycp2, Sycp3, Mak, Tnp1, Tnp2, Odf1, Odf2, and Odf3) are indeed detected only in these cells and not in any of the other cell populations or the controls (see GermOnline for expression patterns). It should be emphasized that this is true for both replicate samples that were analyzed for each cell type. One should bear in mind, however, that for rare cases of genes that are extremely highly expressed in certain (germ) cell types, it may be difficult to distinguish between transcription in the enriched cell population and the subpopulation of contaminating cells (Table 1).

Germ Cell Expression and Function

We included brain and skeletal muscle samples in a refined analysis of 617 probe sets that yield highly expressed in germ cells to identify 121 transcripts and ESTs as being expressed in meiotic and postmeiotic germ cells but not in

mitotic germ cells and at least three different types of somatic controls. This group includes 40 genes whose products are thought to fulfill meiotic or postmeiotic roles. It therefore seems reasonable to assume that 37 ESTs also identified are likely to lead to a number of genes important for spermatogenesis and fertility (Figure 6). It would be very interesting to compare gene expression in testicular tissue with a large number of somatic tissues in the rat to address the important question of germ cell specificity. However, a truly informative tissue profiling approach requires comprehensive data sets that are currently available only for mouse and human (Su *et al*., 2002). The outcome of such an experiment is bound to be extremely yielding in the light of the fact that up to ~4% of the genes in the mouse genome may be expressed in postmeiotic male germ cells alone (Schultz *et al*., 2003).

The work reported here clearly indicates that expression profiling of mammalian spermatogenesis by using highly enriched Sertoli and germ cells as well as testicular and somatic control tissues is feasible and very yielding. The results furthermore confirm that expression data obtained with classical molecular and cell biological methods are faithfully reproduced by microarrays. This exploratory study provides data for numerous genes previously characterized in somatic tissues (and often thought to be specific for them) that indicate expression in testicular germ cells. In addition to that it identifies >290 uncharacterized transcripts potentially important for male fertility. Such results, conveniently accessible through GermOnline and reXbase, will ultimately help better understand the genetic defects that lead to human reproductive pathologies such as sterility and birth defects. Finally, it is likely to contribute to the development of innovative approaches to reversible inhibition of male fertility.

ACKNOWLEDGMENTS

We thank R. Jenni, K. Potter, R. Pöhlmann, and D. Flanders (FMI), for excellent IT infrastructure support within the Basel Computational Biology Center. We thank I. Coiffec and M. Docquier for excellent technical assistance, V. Vallet-Erdtmann for helpful advice, and R. Lindberg and M. Rüegg for having provided unpublished U34A brain and skeletal muscle microarray expression data. P.D. and the GeneChip platform at the Life Sciences Training Facility are supported by the Biozentrum, the Swiss Tropical Institute, the Department of Clinical and Biological Sciences of the Basel University Hospital, and the Department of Clinical Research at the University of Bern. U.S., C.W. and L.H. are supported by a grant from the Swiss Institute of Bioinformatics.

REFERENCES

Adham, I.M., Nayernia, K., Burkhardt-Gottges, E., Topaloglu, O., Dixkens, C., Holstein, A.F., and Engel, W. (2001). Teratozoospermia in mice lacking the transition protein 2 (Tnp2). Mol. Hum. Reprod. *7*, 513–520.

Adham, I.M., Nayernia, K., and Engel, W. (1997). Spermatozoa lacking acrosin protein show delayed fertilization. Reprod. Dev. *46*, 370–376.

Akhtar, A., and Becker, P.B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in *Drosophila*. Mol. Cell *5*, 367–375.

Behrens, A., Genoud, N., Naumann, H., Rulicke, T., Janett, F., Heppner, F.L., Ledermann, B., and Aguzzi, A. (2002). Absence of the prion protein homologue Doppel causes male sterility. EMBO J. *21*, 3652–3658.

Bolstad, B., Irizarry, R., Astrand, M., and Speed, T. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185–193.

Bowdish, K.S., Yuan, H.E., and Mitchell, A.P. (1994). Analysis of RIM11, a yeast protein kinase that phosphorylates the meiotic activator IME1. Mol. Cell Biol. *14*, 7909–7919.

Cho, C., Willis, W.D., Goulding, E.H., Jung-Ha, H., Choi, Y.C., Hecht, N.B., and Eddy, E.M. (2001). Haploinsufficiency of protamine-1 or -2 causes infertility in mice. Nat. Genet. *28*, 82–86.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O., and Herskowitz, I. (1998). The transcriptional program of sporulation in budding yeast. Science *282*, 699–705.

Clarke, G., and Cooke, D. (1998). A Basic Course in Statistics. Oxford, NY: Oxford University Press.

Coaiacovo, M.P., S. G., Reddy, K.C., Reinke, V., Kim, S.K., and Villeneuve, A.M. (2002). A targeted RNAi screen for genes involved in chromosome morphogenesis and nuclear organization in the *Caenorhabditis elegans* germline. Genetics *162*, 113–128.

Cunningham, D.B., Segretain, D., Arnaud, D., Rogner, U., and Avner, P. (1998). The mouse Tsx gene is expressed in Sertoli cells of the adult testis and transiently in premeiotic germ cells during puberty. Dev. Biol. *204*, 345–360.

Deutschbauer, A., Williams, R., Chu, A.M., and Davis, R. (2002). Parallel phenotypic analysis of sporulation and postgermination growth in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA *99*, 15530–15535.

Eddy, E. (2002). Male Germ Cell Gene Expression. Recent Prog. Horm. Res. *57*, 103–128.

Edelman, W., Cohen, P., Kneitz, B., Winand, N., Lia, M., Heyer, J., Kolodner, R., Pollard, J., and Kucherlapati, R. (1999). Mammalian MutS homologue 5 is required for chromosome pairing in meiosis. Nat. Genet. *21*, 123–127.

Freeman, L., Aragon-Alcaide, L., and Strunnikov, A. (2000). The condensin complex governs chromosome condensation and mitotic transmission of rDNA. J. Cell Biol. *149*, 811–824.

Goepfert, T.M., Adigun, Y.E., Zhong, L., Gay, J., Medina, D., and Brinkley, W.R. (2002). Centrosome amplification and overexpression of aurora A are early events in rat mammary carcinogenesis. Cancer Res. *62*, 4115–4122.

Griswold, M. (1998). The central role of Sertoli cells in spermatogenesis. Semin. Cell Dev. Biol. *9*, 411–416.

Gromoll, J., Wessels, J., Rosiepen, G., Brinkworth, M.H., and Weinbauer, G.F. (1997). Expression of mitotic cyclin B1 is not confined to proliferating cells in the rat testis. Biol. Reprod. *57*, 1312–1319.

Guillaume, E., Evrard, B., Com, E., Moertz, E., Jegou, B., and Pineau, C. (2001). Proteome analysis of rat spermatogonia: reinvestigation of stathmin spatiotemporal expression within the testis. Reprod. Dev. *60*, 439–445.

Guo, T., Chan, K., Hakovirta, H., Xiao, Y., Toppari, J., Mitchell, A., and Salameh, W. (2003). Evidence for a role of glycogen synthase kinase-3beta in rodent spermatogenesis. J. Androl. *24*, 332–342.

Irizarry, R., Bolstad, B., Collin, F., Cope, L., Hobbs, B., and Speed, T. (2003a). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. *31*, e15.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003b). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249–264.

Ishibashi, K., Kuwahara, M., Kageyama, Y., Tohsaka, A., Marumo, F., and Sasaki, S. (1997). Cloning and functional expression of a second new aquaporin abundantly expressed in testis. Biochem. Biophys. Res. Commun. *237*, 714–718.

Ishibashi, K., Morinaga, T., Kuwahara, M., Sasaki, S., and Imai, M. (2002). Cloning and identification of a new member of water channel (AQP10) as an aquaglyceroporin. Biochim. Biophys. Acta *1576*, 335–340.

Jégou, B., Pineau, C., and Dupaix, A. (1999). Male Reproductive Function. Boston, MA: Kluwer Academic Publishers.

Kanatsu-Shinohara, M., Toyokuni, S., and Shinohara, T. (2003). CD9 is a surface marker on mouse and rat male germline stem cells. Biol. Reprod.

Kaufman, L., and Rousseeuw, P. (1990). Finding Groups in Data, An Introduction to Cluster Analysis. Brussels, Belgium: John Wiley & Sons.

Keeney, S., Giroux, C.N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell *88*, 375–384.

Kneitz, B., Cohen, P.E., Avdievich, E., Zhu, L., Kane, M.F., Hou, H., Jr., Kolodner, R.D., Kucherlapati, R., Pollard, J.W., and Edelmann, W. (2000). MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. Genes Dev. *14*, 1085–1097.

Le Goffic, R., Mouchel, T., Aubry, F., Patard, J.J., Ruffault, A., Jegou, B., and Samson, M. (2002). Production of the chemokines monocyte chemotactic protein-1, regulated on activation normal T cell expressed and secreted protein, growth-related oncogene, and interferon-gamma-inducible protein-10 is induced by the Sendai virus in human and rat testicular cells. Endocrinology *143*, 1434–1440.

Le Naour, F., Rubinstein, E., Jasmin, C., Prenant, M., and Boucheix, C. (2000). Severely reduced female fertility in CD9-deficient mice. Science *287*, 319–321.

Lipkin, S., *et al*. (2002). Meiotic arrest and aneuploidy in MLH3-deficient mice. Nature Genet. *31*, 385–390.

Liu, D., Matzuk, M., Sung, W., Guo, Q., Wang, P., and Wolgemuth, D. (1998). Cyclin A1 is required for meiosis in the male mouse. Nat. Genet *20*, 377–380.

Livera, G., Rouiller-Fabre, V., Pairault, C., Levacher, C., and Habert, R. (2002). Regulation and perturbation of testicular functions by vitamin A. Reproduction *124*, 173–180.

Lui, W.Y., Lee, W.M., and Cheng, C.Y. (2003). Rho GTPases and spermatogenesis. Biochim. Biophys. Acta *1593*, 121–129.

Luo, G., Yao, M.S., Bender, C.F., Mills, M., Bladl, A.R., Bradley, A., and Petrini, J.H. (1999). Disruption of mRad50 causes embryonic stem cell lethality, abnormal embryonic development, and sensitivity to ionizing radiation. Proc. Natl. Acad. Sci. USA *96*, 7376–7381.

Malathi, K., Xiao, Y., and Mitchell, A.P. (1997). Interaction of yeast repressor-activator protein Ume6p with glycogen synthase kinase 3 homolog Rim11p. Mol. Cell. Biol. *17*, 7230–7236.

Mannan, A.U., Nayernia, K., Mueller, C., Burfeind, P., Adham, I.M., and Engel, W. (2003). Male mice lacking the Theg (testicular haploid expressed gene) protein undergo normal spermatogenesis and are fertile. Biol. Reprod. *69*, 788–796.

McLean, D., Friel, P., Pouchnik, D., and Griswold, M. (2002). Oligonucleotide microarray analysis of gene expression in follicle-stimulating hormone-treated rat Sertoli cells. Mol. Endocrinol. *16*, 2780–2792.

Miki, K., Willis, W.D., Brown, P.R., Goulding, E.H., Fulcher, K.D., and Eddy, E.M. (2002). Targeted disruption of the Akap4 gene causes defects in sperm flagellum and motility. Dev. Biol. *248*, 331–342.

Offenberg, H.H., Schalk, J.A., Meuwissen, R.L., van Aalderen, M., Kester, H.A., Dietrich, A.J., and Heyting, C. (1998). SCP *2*, a major protein component of the axial elements of synaptonemal complexes of the rat. Nucleic Acids Res. *26*, 2572–2579.

O'Shaughnessy, P., Willerton, L., and Baker, P. (2002). Changes in Leydig cell gene expression during development in the mouse. Biol. Reprod. *66*, 966–975.

Ostermeier, G., Dix, D., Miller, D., Khatri, P., and Krawetz, S. (2002). Spermatozoal RNA profiles of normal fertile men. Lancet *360*, 772–777.

Ouspenski, I.I., Cabello, O.A., and Brinkley, B.R. (2000). Chromosome condensation factor Brn1p is required for chromatid separation in mitosis. Mol. Biol. Cell *11*, 1305–1313.

Pang, A.L., *et al*. (2003). Identification of differentially expressed genes in mouse spermatogenesis. J. Androl. *24*, 899–911.

Payne, A. (1990). Hormonal regulation of cytochrome P450 enzymes, cholesterol side-chain cleavage and 17 alpha-hydroxylase/C17–20 lyase in Leydig cells. Biol. Reprod. *42*, 399–404.

Persengiev, S.P., Robert, S., and Kilpatrick, D.L. (1996). Transcription of the TATA binding protein gene is highly up-regulated during spermatogenesis. Mol. Endocrinol. *10*, 742–747.

Petronczki, M., Siomos, M., and Nasmyth, K. (2003). Un menage a quatre: the molecular biology of chromosome segregation in meiosis. Cell *112*, 423–440.

Pineau, C., Syed, V., Bardin, C., Jegou, B., and Cheng, C. (1993a). Germ cell-conditioned medium contains multiple factors that modulate the secretion of testins, clusterin, and transferrin by Sertoli cells. J. Androl. *14*.

Pineau, C., Syed, V., Bardin, C.W., Jegou, B., and Cheng, C.Y. (1993b). Identification and partial purification of a germ cell factor that stimulates transferrin secretion by Sertoli cells. Recent Prog. Horm. Res. *48*, 539–542.

Primig, M., *et al*. (2003). GermOnline: a novel cross-species community annotation database on germ line development and gametogenesis. Nat Genet *35*, 291–292.

Primig, M., Williams, R., Winzeler, E., Tevzadze, G., Conway, A., Hwang, S., Davis, R., and Esposito, R. (2000). The core meiotic transcriptome in budding yeasts. Nat. Genet. *26*, 415–424.

Rabitsch, K., *et al*. (2001). A screen for genes required for meiosis and spore formation based on whole genome expression. Curr. Biol. *11*, 1001–1009.

Reinke, V., *et al*. (2000). A global profile of germline gene expression in *C. elegans*. Mol. Cell *6*, 605–616.

Romanienko, P., and Camerini-Otero, R. (2000). The mouse Spo11 gene is required for meiotic chromosome synapsis. Mol. Cell *6*, 975–987.

Rosok, O., Pedeutour, F., Ree, A.H., and Aasheim, H.C. (1999). Identification and characterization of TESK2, a novel member of the LIMK/TESK family of protein kinases, predominantly expressed in testis. Genomics *61*, 44–54.

Sadate-Ngatchou, P.I., Pouchnik, D.J., and Griswold, M.D. (2003). Identification of testosterone regulated genes in testes of hypogonadal mice using oligonucleotide microarray. Mol. Endocrinol. *18*, 422–433.

Schalk, J.A., Dietrich, A.J., Vink, A.C., Offenberg, H.H., van Aalderen, M., and Heyting, C. (1998). Localization of SCP2 and SCP3 protein molecules within synaptonemal complexes of the rat. Chromosoma *107*, 540–548.

Schlecht, U., and Primig, M. (2003). Mining meiosis and gametogenesis with DNA microarrays. Reproduction *125*, 447–456.

Schultz, N., Hamra, F.K., and Garbers, D.L. (2003). A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. Proc. Natl. Acad. Sci. USA *100*, 12201–12206.

Sha, J., Zhou, Z., Li, J., Yin, L., Yang, H., Hu, G., Luo, M., Chan, H.C., and Zhou, K. (2002). Identification of testis development and spermatogenesis-related genes in human and mouse testes using cDNA arrays. Mol. Hum. Reprod. *8*, 511–517.

Shamsadin, R., Adham, I.M., Nayernia, K., Heinlein, U.A., Oberwinkler, H., and Engel, W. (1999). Male mice deficient for germ-cell cyritestin are infertile. Biol. Reprod. *61*, 1445–1451.

Shinkai, Y., Satoh, H., Takeda, N., Fukuda, M., Chiba, E., Kato, T., Kuramochi, T., and Araki, Y. (2002). A testicular germ cell-associated serine-threonine kinase, MAK, is dispensable for sperm formation. Mol. Cell. Biol. *22*, 3276–3280.

Sicinski, P., *et al*. (1996). Cyclin D2 is an FSH-responsive gene involved in gonadal cell proliferation and oncogenesis. Nature *384*, 470–474.

Smith, E.R., Eisen, A., Gu, W., Sattah, M., Pannuti, A., Zhou, J., Cook, R.G., Lucchesi, J.C., and Allis, C.D. (1998). ESA1 is a histone acetyltransferase that is essential for growth in yeast. Proc. Natl. Acad. Sci. USA *95*, 3561–3565.

Steger, K. (2001). Haploid spermatids exhibit translationally repressed mRNAs. Anat. Embryol. *203*, 323–334.

Storey, J. (2002). A direct approach to false discovery rates. J. R. Stat. Soc. Ser. B, *64*, 479–498.

Su, A., *et al*. (2002). Large-scale analysis of the human and mouse transcriptomes. Proc. Natl. Acad. Sci. USA *99*, 4465–4470.

Toebosch, A., Robertson, D., Klaij, I., de Jong, F., and Grootegoed, J. (1989). Effects of FSH and testosterone on highly purified rat Sertoli cells: inhibin alpha-subunit mRNA expression and inhibin secretion are enhanced by FSH but not by testosterone. J. Endocrinol. *122*, 757–762.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. *3*, Research0034.

Waterston, R.H., *et al*. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Wiederkehr, C., Basavaraj, R., Sarrauste de Menthiere, C., Hermida, L., Koch, R., Schlecht, U., Amon, A., Brachat, S., Breitenbach, M., and Briza, P. (2004a). GermOnline, a cross-species community knowledgebase on germ cell growth and development. Nucleic Acids Res. *32*, D560–D567.

Wiederkehr, C., *et al*. (2004b). Database model and specification of GermOnline Release 2.0, a cross-species community annotation knowledgebase on germ cell differentiation. Bioinformatics (*in press*).

Wolgemuth, D., Laurion, E., and Lele, K. (2002). Regulation of the mitotic and meiotic cell cycles in the male germ line. Recent Prog. Horm. Res. *57*, 75–101.

Yu, Y.E., Zhang, Y., Unni, E., Shirley, C.R., Deng, J.M., Russell, L.D., Weil, M.M., Behringer, R.R., and Meistrich, M.L. (2000). Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. Proc. Natl. Acad. Sci. USA *97*, 4683–4688.

Zickler, D., and Kleckner, N. (1998). The leptotene-zygotene transition of meiosis. Ann. Rev. Genet. *32*, 619–697.